Fairness with an Honest Minority and a Rational Majority

Shien Jin Ong* David Parkes* Alon Rosen* Salil Vadhan*

April 20, 2007

Abstract

We provide a simple protocol for reconstructing the secret in a secret sharing scheme, and prove that it is *fair* when executed with many rational parties together with a small minority of honest parties. That is, when all parties follow the protocol, everyone obtains the secret, and the rational parties have no incentive to deviate from the protocol (because it is a subgame-perfect coalition-proof Nash equilibrium under natural assumptions on the utility functions of the rational parties). The protocol only requires a standard (synchronous) broadcast channel, and tolerates fail-stop deviations from the protocol (i.e. early aborts, but not incorrectly computed messages).

Previous protocols for this problem in the cryptographic or economic models have either required an honest majority, used strong communication channels that enable simultaneous exchange of information, or settled for approximate notions of fairness.

[1]SEAS, Harvard, Cambridge, MA. E-Mail: {shienjin, parkes, alon, salil}@eecs.harvard.edu

1 Introduction

A major concern in the design of distributed protocols is the possibility that parties may deviate from the protocol. Historically, there have been two main paradigms for modeling this possibility. One is the cryptographic paradigm, where some parties are honest, meaning they will always follow the specified protocol, and others are malicious, meaning they can deviate arbitrarily from the protocol. The other is the economic paradigm, where all parties are considered to be rational, meaning that they will deviate from the protocol if and only if it is in their interest to do so.

Recently, some researchers have proposed studying mixtures of these traditional cryptographic and economic models, with various combinations of honest, malicious, and rational participants. One motivation for this that it may allow a more accurate modeling of the diversity of participants in real-life executions of protocols. Along these lines, the papers of Aiyer et al. [3], Lysyanskaya and Triandopoulos [28], and Abraham et al. [2] construct protocols that achieve the best of both worlds. Specifically, they take protocol properties that are known to be achievable in both the cryptographic model (with honest and malicious parties) and the economic model (with only rational parties), and show that protocols with the same properties can still be achieved in a more general model consisting of malicious and rational parties.

Our work is of the opposite flavor. We consider properties that are not achievable in either the cryptographic or economic models alone, and show that they can be achieved in a model consisting of both honest and rational parties. Specifically, we consider the task of secret reconstruction in *secret sharing*, and provide a protocol that is *fair*, meaning that all parties will receive the output, given rational participants together with a small minority of honest participants. In standard communication models, fairness is impossible in a purely economic model (with only rational participants) [22, 24] or in a purely cryptographic model with a small minority of honest participants (and the rest malicious). Previous works achieved fairness by assuming strong communication primitives that allow simultaneous exchange of information [22, 20, 2, 24, 26, 27, 23][1] or settled for approximate notions of fairness [13, 8, 18, 33, 24], whereas we only use a standard (i.e. synchronous but not simultaneous) broadcast channel and achieve the standard notion of fairness.

Thus, our work illustrates the power of a small number of honest parties to maintain equilibria in protocols. These parties follow the specified strategy even when it is not in their interest to do so, whether out of altruism or laziness. Actually, it turns out that, in our protocol, such moves are almost never required during equilibrium play, but instead serve as a threat to keep rational players from deviating. It is the presence of honest players that aligns the incentives of rational players with faithfully following the protocol. We remark that we study our protocol in a simplified model (of only "fail-stop" deviations) that seems to retain the essence of the issues at hand. This enables us to obtain a particularly clean protocol with a clear intuition, which we hope will make it easier for similar ideas to be used in future works handling more complex settings.

Below, we review the cryptographic and economic paradigms in more detail. We then introduce the secret-sharing problem we study and survey recent works on this problem in the purely economic model. We then describe our results and compare them to what was achieved before.

The Cryptographic Paradigm. Here we allow for a subset of the parties to deviate from the protocol in an arbitrary, malicious manner (possibly restricted to computationally feasible strategies), and the actions of these parties are viewed as being controlled by a single adversary.

[1] Actually, the impossibility results of [22, 24] also hold in the presence of a simultaneous broadcast channel and thus the works of [22, 20, 2, 24] use additional relaxations, such as allowing the number of rounds and/or the sizes of the shares to be unbounded random variables.

Intuitively, this captures worst-case deviations from the protocol, so protocols protecting against such malicious and monolithic adversaries provide a very high level of security. Remarkably, this kind of security can be achieved for essentially every multiparty functionality, as shown by a series of beautiful results from the 1980's [39, 19, 9, 6, 34]. However, considering arbitrary (and coordinated) malicious behavior does have some important limitations. For example, it is necessary to either assume that a majority of the participants are honest (i.e. not controlled by the adversary) or allow for protocols that are unfair (i.e. the adversary can prevent some parties from getting the output). This follows from a classic result of Cleve [11], who first showed that there is no fair 2-party protocol for coin-tossing (even with computational security), and then deduced the general version by viewing a multiparty protocol an interaction between two super-parties, each of which controls half of the original parties. Lepinski et al. [26] bypass this impossibility result by assuming a strong communication primitive ("ideal envelopes") which allow simultaneous exchange of information, but it remains of interest to find ways of achieving fairness without changing the communication model.

The Economic Paradigm. Here parties are modeled as rational agents with individual preferences, and will only deviate from the protocol if this is in their own self interest. This approach has become very popular in the computer science literature in recent years, with many beautiful results. There are two aspects of this approach:

1. Design computationally efficient mechanisms (i.e. functionalities that can be implemented by a trusted mediator) that give parties an incentive to be truthful about their private inputs, while optimizing some *social choice function*, which measures the benefit to society and/or the mechanism designer [30, 25, 5].

2. Implement these mechanisms by distributed protocols, with computational efficiency emphasized in *distributed algorithmic mechanism design* [14, 15, 16] and extended to also emphasize additional equilibrium considerations in *distributed implementation* [36, 31, 32], so that parties are "faithful" and choose to reveal private information as well as perform other message passing and computational tasks. More recent works achieve a strong form of distributed implementation, with provably no additional equilibrium [27, 23], but require strong communication primitives and have less focus on computational efficiency.

Note that distributed algorithmic mechanism design is different in spirit from the traditional problem considered in cryptographic protocols, in that parties have "true" private inputs (whereas in cryptography all inputs are considered equally valid) and there is freedom to change how these inputs are mapped to outcomes (whereas in cryptography, the functionality is pre-specified.) Nevertheless, recent works have explored whether we can use the economic model to obtain 'better' solutions to traditionally cryptographic problems, namely to compute some pre-specified functionalities. One potential benefit is that we may be able to incentivize parties to provide their "true" private inputs (whereas in cryptography all inputs are considered equally valid) along the lines of Item 1 above; the papers [29, 37] explore for what functionalities and kinds of utility functions this is possible.

A second potential benefit is that rational deviations may be easier to handle than malicious deviations (thus possibly leading to protocols with better properties), while also preferable to assuming a mixture of players at the honest and malicious extremes. This has led to a line of work, started by Halpern and Teague [22] and followed by [20, 2, 24], studying the problems of secret sharing and multiparty computation in the purely economic model, with all rational participants. One can also require notions of equilibria that are robust against coalitions of rational players [2].

While this approach has proved to be quite fruitful, it too has limitations. Specifically, as pointed out in [20, 24], it seems difficult to construct rational protocols that are fair in the standard communication model, because parties may have an incentive to abort once they receive their own output. The works [22, 20, 2, 24], as well as [27, 23] applied to appropriately designed mediated games, achieve fairness by using strong communication primitives (simultaneous broadcast, "ballot boxes") that allow simultaneous exchange of information.

As mentioned above, we achieve fairness in the standard communication model by considering a mix of rational participants together with a *small* minority of honest participants. Note that Cleve's [11] proof that an honest majority is necessary in the cryptographic setting, by reduction to the two-party case, no longer applies. The reason is that we cannot view a subset of the rational parties as being controlled by a single super-party. Even in coalitional notions of equilibria, each individual in that subset would only agree to a coordinated (joint) deviation if it is in its own interest to do so.

Our protocol is for the share reconstruction problem in secret sharing, which we now describe in more detail.

Secret Sharing. In a *t-out-of-n secret-sharing scheme* [35, 7], a dealer takes a secret s and computes n (randomized) *shares* $s_1, \ldots, s_n$ of s, which are distributed among n parties. The required properties are that (a) any set of t parties can reconstruct the secret s from their shares, but (b) any set of fewer than t parties has no information about s (i.e. they would have been equally likely to receive the same shares for every possible value of s).

Secret sharing is a fundamental building block for cryptographic protocols [19, 6, 9, 34]. Typically, these protocols are structured as follows. First, every party shares its private input among all the parties. Then the computation of the functionality is done on shares (to maintain privacy). And at the end, the parties reveal their shares of the output so that everyone can reconstruct it. Our focus in this paper is on this final reconstruction step. Typically, it is assumed that there are enough honest parties in the protocol to ensure that the secret can be reconstructed from the revealed shares, even if some parties refuse to reveal their shares. A more challenging scenario is one where some parties may reveal incorrect values, which is handled by use of *verifiable secret sharing* [10], but for simplicity in this paper we only consider *fail-stop deviations*, where a party may abort the protocol early but otherwise follows the prescribed strategy. (For example, this models people who may disconnect their computer from the network in the middle of the protocol, but do not have the time or skill to reprogram the software.) If we allow arbitrary fail-stop deviations, then it is clear that having $k \geq t$ honest parties are necessary and sufficient to have a reconstruction protocol that is *fair*, where everyone obtains the secret if anyone does. (In applications of secret sharing to secure multiparty computation, it is typically also important that the threshold is greater than the number of malicious parties, i.e. $t > n - k$. Combined with the previous statement, this implies that there are more honest parties than malicious ones, i.e. we need an honest majority.)

Rational Secret Sharing. It is natural to ask whether we can bypass this lower bound by considering only *rational* deviations from the protocol. The study of secret sharing with only rational participants was initiated by Halpern and Teague [22], and there have been several subsequent works [20, 24, 2]. In these works, it is assumed that participants prefer to learn the secret over not learning the secret, and secondarily, prefer that as few other agents as possible learn it. As pointed out in [20], any protocol where rational participants reveal their shares sequentially will not yield a Nash equilibrium. It is rational for the t'th player to abort, as she can already compute the secret from the shares of the first $t-1$ players and her own, and aborting may prevent the first $t-1$

players from learning it.

One way to get around this difficulty is to assume a *simultaneous broadcast channel*, where all parties can broadcast values at the same time, without the option of waiting to see what values the other parties are broadcasting. All parties simultaneously revealing their shares is a Nash equilibrium. That is, assuming all of the other parties are simultaneously revealing their shares, no party can increase their utility by aborting instead of revealing. However, this basic protocol has several deficiences:

1. A simultaneous broadcast channel is a strong (and perhaps unrealistic) communication primitive, particularly in the context of trying to achieve fairness, where the typical difficulties are due to asymmetries in the times that parties get information. For example, fair coin-tossing is trivial with a simultaneous broadcast channel (everyone broadcasts a bit, and the output is the exclusive-or), in contrast to Cleve's impossibility result for synchronous broadcast channels [11].

2. Nash Equilibrium in this context is a very weak guarantee. For example, as argued by Halpern and Teague [22], it seems likely that rational parties would actually abort. The reason is that aborting is never worse than revealing, and is sometimes better (if $t-1$ other parties reveal, then the tth party will always learn the secret and can prevent the other parties from doing so by an abort.) In addition, plain Nash Equilibrium does not consider deviations by *coalitions* of participants.

Halpern and Teague [22] and follow-up works [20, 2, 24] focus on the second issue. That is, they allow simultaneous broadcast, and explore whether stronger solution concepts than plain Nash equilibrium can be achieved. Halpern and Teague [22] propose looking for an equilibrium that survives "iterated deletion of weakly dominated strategies." They prove that no bounded-round protocol can achieve a fair outcome in equilibrium when adopting this solution concept. However, they and subsequent works by Gordon and Katz [20] and Abraham et al. [2] show that it fair outcomes are possible even with this equilibrium refinement using a probabilistic protocol whose number of rounds has finite expectation. Moreover, Abraham et al. [2] shows how to achieve an equilibrium that is resistant to deviations by coalitions of limited size. Kol and Naor [24] argue that "strict equilibria" is a preferable solution concept to the iterated deletion notion used by [22], and show how to achieve it with a protocol where the size of shares dealt is an unbounded random variable with finite expectation. (They also show that a strict equilibrium cannot be achieved if the shares are of bounded size.) In all of the above works, the protocols' prescribed instructions crucially depend on the utilities of the various players.

The works of Lepinski et al. [27] and Izmalkov et al. [23] also can be used to obtain fair protocols for secret sharing by making an even stronger physical assumption than a simultaneous broadcast channel, namely "ballot boxes." Specifically, they show how to compile any game with a trusted mediator into a fair ballot-box protocol with the same incentive structure. Since the share-reconstruction problem has a simple fair solution with a trusted mediator (the mediator takes all the inputs, and broadcasts the secret iff *all* players reveal their share), we can apply their compiler to obtain a fair ballot-box protocol.

Our Results. We focus primarily on the first issue: our goal is to achieve fairness without a simultaneous broadcast channel, but only a *synchronous broadcast channel*. That is, the protocol should proceed in rounds, and only one party can broadcast in each round.[2] When all parties are

[2]For round efficiency, sometimes people use a slightly more general channel where many parties can broadcast in a single round, but deviating parties are can perform 'rushing' — wait to see what others have broadcast before

rational, the only positive result is in independent work by Kol and Naor [24], who achieve an approximately fair solution (specifically, a fair ε-Nash equilibrium). We instead assume that there is a *small* number k of honest participants (which can be much smaller than the secret-sharing threshold t), and the rest are rational. Our main result is that in this setting, there is a simple protocol that achieves fair outcomes in a very robust equilibrium refinement.

Our protocol is simple to describe. We assume that there is a random ordering π of the participants, chosen and published at the start of the protocol, say by the dealer of secret-sharing scheme. Then the participants proceed according to this ordering, broadcasting their shares sequentially. (There are some cases that participants are instructed to abort, if certain earlier participants aborted.)

If all parties follow the specified strategy, then everyone will reveal their share and we have a fair outcome. We show that with very high probability, namely $1 - (t/n)^k$, over the permutation π, the specified strategy is an equilibrium. It is not just a plain Nash equilibrium, but achieves a number of stronger and attractive solution concepts. First, it is *subgame perfect*, which means that the strategy is rational to follow regardless of the previous history of messages; intuitively, this means that the equilibrium does not rely on irrational empty threats (where one player will punish another player for deviating even at his own expense). Second, it is *coalition-proof* in the sense of Bernheim, Peleg, and Whinston [4]; intuitively, this means that no coalition of players can deviate in a way that is stable and (Pareto) improves their utilities. As mentioned above, the previous works [22, 20, 2, 24] also use strengthenings of Nash equilibria (iterated admissible equilibria, k-resilient equilibria, strict equilibria). Those notions are incomparable to ours, but our point is to achieve results without using a simultaneous broadcast channel. In addition, our protocol is significantly simpler than the previous ones and is independent of the numeric values in the utility functions of the players, in contrast to [22, 20, 2, 24]. In this sense, the protocol is "detail-free" in the sense of Wilson's critique [38] of approaches in mechanism design that require detailed information about the preferences of parties and achieves additional robustness. Our protocol also has a bounded number of rounds and does not require changing the underlying secret-sharing scheme.

Notice that if $t = \alpha n$ for some constant $\alpha \in (0, 1)$ and $k = \omega(\log n) \ll t$, we achieve an equilibrium with all but negligible $(1/n^{\omega(1)})$ probability over the permutation π. In contrast, if we had malicious participants rather than rational ones, then we would need $k \geq t$ honest parties to have a fair solution where everyone learns the secret. On the other hand, there is no fair solution with only rational parties. Thus, by considering a mixture of rational and honest participants, we achieve something that is impossible in either the purely cryptographic or purely economic frameworks.

2 The Protocol

In this section we define the game-theoretic setting in which the secret-sharing reconstruction protocol is performed, and describe the prescribed strategy that all parties will be asked to follow (whether rational or honest). We also define our model for the preferences of rational parties for different outcomes of the protocol and make precise our communication model. We delay making explicit our particular rational/honest mixture model until Section 3.

broadcasting their own values.

2.1 Our Protocol Setting

Here we informally describe the setting of our protocol, which we will subsequently formalize as a game with a specified strategy.

Recall the informal description of secret sharing from the Introduction. We assume an honest dealer D holding a (uniformly random) secret σ, and n players $1, \ldots, n$. There is also a threshold $t \leq n$, known to all players, which is fixed at the outset. At the beginning of the protocol, D computes *shares* $s_1, \ldots, s_n$ of the secret s, and distributed share s_i to player i. The properties of a secret-sharing scheme tell us that no subset of fewer than t players should have any information about σ at this point. The dealer also distributes a uniformly chosen permutation $\pi \in S_n$. ¿From that point on, the dealer does not take part in the protocol.

Now, our protocol proceeds in a sequence of rounds, where in each round a single player can broadcast her share to all other parties (using a *synchronous broadcast* channel). The order in which the players proceed is determined by the permutation π. Specifically, at round ℓ of the protocol, it is the turn of player $i = \pi(\ell)$ to broadcast. She can either *reveal* her share s_i or *abort* the protocol. The specified strategy of our protocol will require that she reveals s_i unless one of the first $t-1$ parties to speak (i.e. parties $\pi(1), \ldots, \pi(t-1)$) has aborted, in which case she should also abort.

We will consider *fail-stop* deviations from this protocol, which means that a party may abort the protocol even when the specified strategy says that she should reveal her share. That is, parties cannot broadcast out of turn or send values other than their correct share. We do *not* banish (or punish) a party that aborts; that is, the party continues to hear subsequent broadcasts on the channel. By the properties of secret sharing, it follows that a party will be able to compute the secret at the end of the protocol if at least $t-1$ *other* parties have revealed their shares, and otherwise she has no information about the secret.

2.2 Formalization as a Game

We now abstract the protocol described in the previous section as a game with a specified strategy profile.

Definition 2.1 (Secret Sharing Game)**.** *Let $\pi \in S_n$ be a permutation. The* secret sharing game *is defined as the extensive form game $\Gamma^\pi = (N, H, P^\pi, u)$, where:*

- *The set of players is $N = \{1, \ldots, n\}$,*
- *The set of histories is $H = \bigcup_{\ell=0}^{n} \{\texttt{REVEAL}, \texttt{ABORT}\}^\ell$,*
- *for any $h \in H$ of length $\ell - 1$, the next message function satisfies $P^\pi(h) = \pi(\ell)$, and*
- *the set Z of terminal histories consists of all $h = (a_1, \ldots, a_n)$ of length n.*

For any fixed π the game tree of Γ^π is a full binary tree of depth n, where the ℓ^{th} level of the tree fully corresponds to player i where $i = \pi(\ell)$. In other words, the only level in which player i takes action is level $\ell = \pi^{-1}(i)$, and player i is the only player to take action in this level. Note that each game Γ^π refers to a fixed permutation π. The reason that the dealer chooses the permutation at random is that our results will only establish equilibrium properties of Γ^π for *most* choices of π.

A protocol is defined through a *prescribed strategy* $\mathbf{s} = (s_1, \ldots, s_n)$ for the game Γ^π. If for all $i \in N$, player i follows strategy s_i, the vector $\mathbf{s}$ will represent a joint execution of the protocol by all n players. The value $o(\mathbf{s})$ will then represent the game's outcome, and $u_i(\mathbf{s})$ the actual value of player i's payoff.

Definition 2.2 (Protocol for secret sharing)**.** *A* protocol for secret sharing *is a pair* $(\Gamma^\pi, \mathbf{s})$*, where* Γ^π *is the secret sharing game, and* $\mathbf{s} = (s_1 \ldots, s_n)$ *is a strategy vector for* Γ^π*.*

We now formally describe the prescribed strategy corresponding to our protocol. Our ultimate objective is to demonstrate that this strategy achieves various notions of equilibria (given natural assumptions on the utilities of the players) and thus we can expect that it will be followed by rational players.

Definition 2.3 (The Prescribed strategy)**.** *The* prescribed strategy *for* Γ^π *is the strategy vector* $\mathbf{s} = (s_1 \ldots, s_n)$ *defined as follows. For* $i \in N$*, the input to the strategy function* s_i *is a non-terminal history* $h = (a_1, \ldots, a_{\ell-1}) \in H \backslash Z$*, where* $i = \pi(\ell)$*. In accordance with the definition of the game* Γ^π*, we will have that the next player* $P^\pi(a_1, \ldots, a_{\ell-1})$ *to take action is player* i*. Player* i*'s prescribed strategy is defined as follows.*

$$s_i(a_1, \ldots, a_{\ell-1}) = \begin{cases} \texttt{REVEAL} & \text{if } a_j = \texttt{REVEAL} \text{ for all } \quad j < \min\{\ell, t\} \\ \texttt{ABORT} & \text{if } a_j = \texttt{ABORT} \text{ for some } j < \min\{\ell, t\} \end{cases} \tag{1}$$

Note that if all players follow $\mathbf{s}$ then all n players (and in particular at least t out of the n players) eventually reveal their share. Thus, if we can argue that it is rational for all players to follow $\mathbf{s}$ we will have obtained a fair solution to the secret sharing problem (since all players learn the secret following the protocol's execution).

A crucial property of the prescribed strategy is that *even players that adhere to it do not necessarily reveal their share.* This happens as soon as one of the first $t-1$ players aborts. On the other hand, if none of the first $t-1$ players aborts, the prescribed strategy instructs to reveal, even if the subsequent history (i.e., somewhere after the first $t-1$ rounds) does contain an abort.

2.3 Assumptions on the Utility Functions

We now describe our assumptions regarding the utility functions of the rational players. For a particular history $h \in H$ in the game Γ^π, we let $\delta_i(h)$ be a Boolean value denoting whether or not player i learns the secret from the history so far. In particular, given a specific outcome o of the game (which is merely a terminal history $o \in Z$), the bit $\delta_i(o)$ indicates whether or not i learns the secret from the execution of the protocol. That is, $\delta_i(o) = 1$ iff at least $t-1$ parties other than i revealed during the protocol. Let $\mathsf{num}(o) = \sum_i \delta_i(o)$; i.e., $\mathsf{num}(o)$ is simply the total number of players who learn the secret from the protocol's execution. Let $u_i(o)$ denote the utility of player i for the outcome o. Following [22, 20], we make the following assumptions about the utility functions of the players:

1. $\delta_i(o) > \delta_i(o') \Rightarrow u_i(o) > u_i(o')$.
2. If $\delta_i(o) = \delta_i(o')$, then $\mathsf{num}(o) < \mathsf{num}(o') \Rightarrow u_i(o) > u_i(o')$.

That is, player i first prefers outcomes in which he learns the secret; as long as δ_i remains constant, player i prefers strategies in which the fewest number of other players learn the secret. These ordinal utility function assumptions will be sufficient for the analysis of the equilibrium properties of the protocol. In fact, our analysis will in fact hold even if we replace condition (2) above with a somewhat weaker condition:

3. If $\delta_i(o) = \delta_i(o')$, then $\mathsf{learn}(o) \subsetneq \mathsf{learn}(o') \Rightarrow u_i(o) > u_i(o')$,

where for an outcome o, we define $\mathsf{learn}(o)$ to be the set of players $j \in N$ for which $\delta_j(o) = 1$. In other words, as long as δ_i remains constant, player i has a preference of an outcome o over an outcome o' only if the set of players that learn the secret in o is *strictly contained* in the set of players that learn the secret in o'.[3]

2.4 Comments

Remark 2.1. *Previous solutions [22, 20] crucially rely on a* simultaneous broadcast *channel. That is, for their analysis to go through, it is necessary to assume that all parties broadcast their value* before *they have received any value that was broadcast by other parties (i.e. no "rushing" is allowed). In the context of fairness of protocols, assuming "plain" broadcast seems much weaker than assuming simultaneous broadcast (since the latter seems much "closer" to the solution of the problem.).*

3 Introducing an Honest Minority

Recall that our goal is to argue that: (1) it is rational for the players to follow the prescribed strategy, and (2) if everybody follows the prescribed strategy at least t shares are revealed. (Note that revealing only $t-1$ shares will not be sufficient because the parties that revealed one of these shares will not themselves learn the secret.)

3.1 Impossibility if Everybody is Rational

The setting described in Section 2 is quite demanding and makes the task of designing fair protocols challenging, There are two main sources of difficulty: (1) The only communication channel allowed to the players is synchronous broadcast, and (2) the number of communication rounds in the protocol is a-priori bounded by n (as every player is allowed to broadcast only once).

Here we argue that, in the above setting, any protocol for secret sharing for which the prescribed strategy s is a subgame perfect equilibrium for Γ^π, the outcome $o = o(s)$ of the protocol (Γ^π, s) necessarily satisfies $\mathsf{num}(o) = 0$. While, strictly speaking, this is not considered a violation of fairness (since nobody eventually learns the secret), it still implies that no satisfactory solution would be possible in this setting (though alternative settings are not inconceivable). Thus, in order to obtain a meaningful result in the context of fairness, some modification will be necessary. As we have already mentioned, our way to bypass the impossibility result will be to assume that a small subset of honest players honestly follows the prescribed strategy.

Proposition 3.1. *Let s be a subgame perfect equilibrium for the game Γ^π, Then, $\mathsf{num}(o) = 0$.*

3.2 Restricted Games

To get around the above impossibility result, we will assume that a small subset of *honest* players always follows the prescribed strategy (whether or not this is the best response to other players' actions). In order to formalize this situation, we introduce the notion of a "restricted game" for the remaining rational players. Restricted games will also be useful at a later stage in defining various notions of coalition proof equilibria. Loosely speaking, a *restricted game* is obtained by fixing the

[3] To see why condition (3) is weaker than (2), observe that whenever $\mathsf{learn}(o)$ is not contained in $\mathsf{learn}(o')$, player i will have no preference, even if $\mathsf{num}(o) \geq \mathsf{num}(o')$.

strategies of some players outside of some subset K to follow a pre-specified strategy vector s. A formal definition can be found in the Appendix.

As mentioned in Section 2.2, the prescribed strategy $\mathbf{s}$ may in some cases instruct an honest player to abort. Allowing honest players to sometimes abort ensures that players who deviate from the protocol's prescribed strategy in early rounds (the first $t-1$) are penalized and unable to learn the secret themselves. On the other hand, the honest players will continue to report their share even if a deviation occurs in the t^{th} round; this ensures that the players in the first $t-1$ rounds will learn the secret in addition to the other players, and thus maintains fairness. It is helpful to think of the honest players as distributing the role of a trusted mediator: they make sure that fairness is guaranteed for all participants and allow for a game-theoretic solution without simultaneous communication.

3.3 Good Permutations

We begin by defining a condition on a permutation π that is sufficient in order to guarantee that $\mathbf{s}$ satisfies our various notions of equilibria for the game $\Gamma^{\pi}_{\mathbf{s}_{-K}}$.

Definition 3.2 (Good permutation)**.** *Let $K \subset N$. A permutation $\pi \in S_n$ is said to be K-*good *if there exist $k \geq t$ so that $\pi(k) \in N \setminus K$. Otherwise, it is said to be K-*bad*.*

In other words, a permutation π is K-good if there exists an honest player whose turn to broadcast occurs after the $(t-1)^{\text{st}}$ round of the game Γ^{π}. We observe that most permutations are good.

Proposition 3.3. *For every set $K \subset N$, the probability that a randomly chosen permutation π is K-bad is at most $(t/n)^{|N \setminus K|}$.*

Thus, as long as the set $N \setminus K$ of honest players is of size at least $\omega((\log n)/\log(n/t))$, a random π will be K-bad with only negligible ($n^{-\omega(1)}$) probability. Thus, unless t is very close to n, a very small minority of honest players suffices. We remark that instead of having the dealer choose the permutation at random, one might instead use a random selection protocol, whereby the participants choose the permutation themselves. In particular, the random selection protocols of [21] only require a minority of honest participants, and will maintain the property that the selected permutation is K-bad with vanishingly small probability.

4 Subgame Perfect Equilibrium

We prove that for any set $K \subset N$ and any K-good permutation π, the prescribed strategy vector $\mathbf{s}_K$ is a subgame perfect equilibrium for the extensive game $\Gamma^{\pi}_{\mathbf{s}_{-K}}$.

Theorem 4.1 (Subgame perfect equilibrium)**.** *Let $K \subset N$ and suppose π is a K-good permutation. Then, the strategy vector $\mathbf{s}_K$ is a subgame perfect equilibrium for the game $\Gamma^{\pi}_{\mathbf{s}_{-K}}$.*

5 Coalition Proof Equilibrium

In this section, we strengthen our result from Section 4 and show that our secret sharing protocol satisfies a stronger notion of equilibrium, namely our prescribed strategy $\mathbf{s}$ is resilient against coalitional deviations, as opposed to just the single player deviations considered in Section 4.

Specifically, we show that our prescribed strategy is a *perfectly coalition-proof Nash equilibrium (PCPNE)* in the sense of Bernheim, Peleg, and Whinston [4]. Informally, a strategy is PCPNE if there is no *self-enforcing* coalition of players that can deviate and gain; by self-enforcing, we mean that there is no subcoalition of players that can redeviate and "gain," in such a way that their redeviation is PCPNE. Observe that the notion of self enforcing and PCPNE is defined recursively. We present a formal definition of PCPNE next.

5.1 Perfectly Coalition-Proof Nash Equilibrium (PCPNE)

First, we define what it means for a coalition to "gain." We use a Pareto improvement notion instead of requiring that all players in the coalition strictly improve their utilities.

Definition 5.1 (Pareto improvement)**.** *For a game $\Gamma = (N, H, P, u)$, strategy s is said to* Pareto-improve *strategy s' with respect to set $C \subseteq N$ if for all $i \in C$, $u_i(s') \geq u_i(s)$, and there exists a $j \in C$ such that $u_j(s') > u_j(s)$.*

Following [4], we define PCPNE for an extensive game inductively on the number of players and number of stages.

Definition 5.2 (perfectly coalition-proof Nash equilibrium, following [4][4])**.** *Strategy s^* is a* perfectly coalition-proof Nash equilibrium (PCPNE) *in a game $\Gamma = (N, H, P, u)$ with n players and r stages, where $(n, r) \neq (1, 1)$, if the following two conditions hold.*

1. *s^* is* perfectly self enforcing (PSE) *in Γ, namely:*
   (a) *for all $C \subsetneq N$, s^*_C is PCPNE in $\Gamma_{s^*_{-C}}$, and*
   (b) *the restriction of s^* to any proper subgame forms a PCPNE in that subgame. In other words, for every history $h \in H \setminus \{\varepsilon\}$, $s^*|_h$ is a PCPNE in $\Gamma|_h$.*
2. *There does not exists another PSE strategy s in Γ that Pareto-improves s^* with respect to N.*

When $(n, r) = (1, 1)$, strategy s^* is a PCPNE (and PSE) in a single-player, single-stage game Γ if $u_1(s^*) \geq u_1(s)$ for all $s \in S$.

Remark 5.1. *Our definition above differs from the definition given by Bernheim, Peleg, and Whinston [4] in that we increase the set of candidate deviations by allowing only Pareto improvement in Item 1b above, whereas [4] require **strict** improvement for all players. Due to the recursive nature of the definition, these two equilibrium concepts—ours and that of [4]—seem incomparable. Requiring weak Pareto improvement, however, is more natural in our setting because we wish to also allow for deviations in which some players may be indifferent.*

PCPNE is a stronger equilibrium concept than the subgame-perfect equilibrium (SPE) concept discussed in Section 4. The converse does not hold in general—there are games with strategies that are SPE but not PCPNE (c.f., [4, Section 4]).

Claim 5.3. *For any game Γ, if s^* is PCPNE in Γ, then s^* is SPE in Γ.*

We now state our main theorem.

Theorem 5.4. *Let protocol $(\Gamma^{\pi}_{\mathbf{s}_{-K}}, \mathbf{s}_K)$ be defined as in Definition 2.2, with π being a K-good permutation and $\mathbf{s}$ being our prescribed strategy. Then, strategy $\mathbf{s}_K$ is PCPNE in the game $\Gamma^{\pi}_{\mathbf{s}_{-K}}$.*

[4]See also Ferreira [17] for a discussion of possible additional considerations that are not incorporated within the Bernheim, Peleg, and Whinston [4] definition. These additional considerations, however, seem irrelevant in our perfect information extensive game in which each players moves at most once.

References

[1] *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing, 2-4 May 1988, Chicago, Illinois, USA*. ACM, 1988.

[2] I. Abraham, D. Dolev, R. Gonen, and J. Y. Halpern. Distributed computing meets game theory: robust mechanisms for rational secret sharing and multiparty computation. In E. Ruppert and D. Malkhi, editors, *PODC*, pages 53–62. ACM, 2006.

[3] A. S. Aiyer, L. Alvisi, A. Clement, M. Dahlin, J.-P. Martin, and C. Porth. Bar fault tolerance for cooperative services. In A. Herbert and K. P. Birman, editors, *SOSP*, pages 45–58. ACM, 2005.

[4] B. P. B. Douglas Bernheim and M. D. Whinston. Coalition-proof nash equilibria i. concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.

[5] M. Babaioff, R. Lavi, and E. Pavlov. Mechanism design for single-value domains. In *Proc. Nat. Conf. on Artificial Intelligence, AAAI05*, 2005.

[6] M. Ben-Or, S. Goldwasser, and A. Wigderson. Completeness theorems for non-cryptographic fault-tolerant distributed computation (extended abstract). In *STOC* [1], pages 1–10.

[7] G. Blakely. Safeguarding cryptographic keys. In *AFIPS*, volume 48, page 313, 1979.

[8] D. Boneh and M. Naor. Timed commitments. In *Proc. CRYPTO 2000*, pages 236–254, 2000.

[9] D. Chaum, C. Crépeau, and I. Damgård. Multiparty unconditionally secure protocols (extended abstract). In *STOC* [1], pages 11–19.

[10] B. Chor, S. Goldwasser, S. Micali, and B. Awerbuch. Verifiable secret sharing and achieving simultaneity in the presence of faults (extended abstract). In *FOCS*, pages 383–395. IEEE, 1985.

[11] R. Cleve. Limits on the security of coin flips when half the processors are faulty (extended abstract). In *STOC*, pages 364–369. ACM, 1986.

[12] C. Dwork, editor. *Advances in Cryptology - CRYPTO 2006, 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006, Proceedings*, volume 4117 of *Lecture Notes in Computer Science*. Springer, 2006.

[13] S. Even, O. Goldreich, and A. Lempel. A randomized protocol for signing contracts. *Commun. ACM*, 28(6):637–647, 1985.

[14] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP-based mechanism for lowest-cost routing. In *Proceedings of the 2002 ACM Symposium on Principles of Distributed Computing*, pages 173–182, 2002.

[15] J. Feigenbaum, C. H. Papadimitriou, and S. Shenker. Sharing the cost of multicast transmissions. *Journal of Computer and System Sciences*, 63:21–41, 2001.

[16] J. Feigenbaum and S. Shenker. Distributed Algorithmic Mechanism Design: Recent Results and Future Directions. In *Proceedings of the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pages 1–13, 2002.

[17] J. Ferreira. A communication-proof equilibrium concept. *Journal of Economic Theory*, 68:249–257, 1996.

[18] J. A. Garay and M. Jakobsson. Timed release of standard digital signatures. In *Proc. Financial Cryptography 2002*, pages 168–182, 2002.

[19] O. Goldreich, S. Micali, and A. Wigderson. How to play any mental game or a completeness theorem for protocols with honest majority. In *STOC*, pages 218–229. ACM, 1987.

[20] S. D. Gordon and J. Katz. Rational secret sharing, revisited. In R. D. Prisco and M. Yung, editors, *SCN*, volume 4116 of *Lecture Notes in Computer Science*, pages 229–241. Springer, 2006.

[21] R. Gradwohl, S. P. Vadhan, and D. Zuckerman. Random selection with an adversarial majority. In Dwork [12], pages 409–426.

[22] J. Y. Halpern and V. Teague. Rational secret sharing and multiparty computation: extended abstract. In L. Babai, editor, *STOC*, pages 623–632. ACM, 2004.

[23] S. Izmalkov, S. Micali, and M. Lepinski. Rational secure computation and ideal mechanism design. In *FOCS*, pages 585–595. IEEE Computer Society, 2005.

[24] G. Kol and M. Naor. Games for exchanging information. manuscript. 2007.

[25] D. Lehmann, L. I. O'Callaghan, and Y. Shoham. Truth revelation in approximately efficient combinatorial auctions. *Journal of the ACM*, 49(5):577–602, September 2002.

[26] M. Lepinski, S. Micali, C. Peikert, and A. Shelat. Completely fair sfe and coalition-safe cheap talk. In S. Chaudhuri and S. Kutten, editors, *PODC*, pages 1–10. ACM, 2004.

[27] M. Lepinski, S. Micali, and A. Shelat. Collusion-free protocols. In H. N. Gabow and R. Fagin, editors, *STOC*, pages 543–552. ACM, 2005.

[28] A. Lysyanskaya and N. Triandopoulos. Rationality and adversarial behavior in multi-party computation. In Dwork [12], pages 180–197.

[29] R. McGrew, R. Porter, and Y. Shoham. Towards a general theory of non-cooperative computation. In J. Y. Halpern and M. Tennenholtz, editors, *TARK*, pages 59–71. ACM, 2003.

[30] N. Nisan and A. Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35:166–196, 2001.

[31] D. C. Parkes and J. Shneidman. Distributed implementations of Vickrey-Clarke-Groves mechanisms. In *Proc. 3rd Int. Joint Conf. on Autonomous Agents and Multi Agent Systems*, pages 261–268, 2004.

[32] A. Petcu, B. Faltings, and D. Parkes. M-dpop: Faithful distributed implementation of efficient social choice problems. In *AAMAS'06 - Autonomous Agents and Multiagent Systems*, pages 1397–1404, Hakodate, Japan, May 2006.

[33] B. Pinkas. Fair secure two-party computation. In *Proc. EUROCRYPT 2003*, pages 87–105, 2003.

[34] T. Rabin and M. Ben-Or. Verifiable secret sharing and multiparty protocols with honest majority (extended abstract). In *STOC*, pages 73–85. ACM, 1989.

[35] A. Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979.

[36] J. Shneidman and D. C. Parkes. Specification faithfulness in networks with rational nodes. In *Proc. 23rd ACM Symp. on Principles of Distributed Computing (PODC'04)*, St. John's, Canada, 2004.

[37] Y. Shoham and M. Tennenholtz. Non-cooperative computation: Boolean functions with correctness and exclusivity. *Theor. Comput. Sci.*, 343(1-2):97–113, 2005.

[38] R. Wilson. Game-theoretic approaches to trading processes. In T. Bewley, editor, *Advances in Economic Theory: Fifth World Congress*. Cambridge University Press, 1987.

[39] A. C.-C. Yao. How to generate and exchange secrets (extended abstract). In *FOCS*, pages 162–167. IEEE, 1986.

A Preliminaries

A.1 Extensive Form Games with Perfect Information

Our protocol will consist of a prescribed strategy vector for an *extensive form game with perfect information*. Considering extensive games allows us to model the *sequential* nature of protocols, where each player considers his plan of action only following some of the other players' actions. The notion of perfect information captures the fact that each player, when making any decision, is perfectly informed of all the events that have previously occurred.

Jumping ahead, we mention that the above choice is consistent with our protocol setting, where players' possible actions in a given round of a protocol amount to either broadcasting their value or refusing to do so. In both cases the choice of the action is visible by all other players.

Definition A.1 (Extensive game with perfect information)**.** *An* extensive form game with perfect information *(or* game *for short) is a tuple* (N, H, P, u) *where:*

- $N = \{1, \ldots, n\}$ *is the set of* players*.*
- H *is a set of (finite)* history *sequences satisfying that the empty word* $\epsilon \in H$*. The components of a history sequence* $h \in H$ *are called* actions*. We let* $A(h) = \{a : (h, a) \in H\}$*. A history* $h \in H$ *is* terminal *if* $A(h) = \emptyset$*. The set of terminal histories is denoted* Z*.*
- $P : (H \setminus Z) \to N$ *is a function that assigns a "next" player to every non-terminal history.*
- $u = (u_1, \ldots, u_n)$ *is a vector of* payoff *functions* $u_i : Z \to \mathbb{R}$*.*

We interpret an extensive form game as follows: after any nonterminal history $h \in H \setminus Z$ player $N(h)$ chooses an action from the set $A(h)$. The empty history $h_0 = (\epsilon)$ is the starting point of the game. Player $P(\epsilon)$ chooses an action $a \in A(\epsilon)$. This induces a history $h_1 = (a)$, and player $P(h_1)$ subsequently chooses an action from the set $A(h_1)$; this choice determines the next player to move, and so on until a terminal history $h_{fin} \in Z$ is reached. The payoff of player i in the execution of the game is then determined to be the value $u_i(h_{fin})$.

An extensive form game can be represented by a tree, where each node of the tree is labeled with a different history $h \in H$. The label of the tree's root, which represents the beginning of

the game, is the empty history $h = \epsilon$. The leaves are labeled with terminal histories $h \in Z$, and correspond to the end of the game. Each internal node of the tree belongs to a player in the sense that it represents a stage in the game in which it is that player's move. Specifically, node h belongs to player $P(h)$, and its outgoing edges correspond to the set $A(h)$. The leaves, which correspond to terminal histories $h \in Z$, are also labeled with a corresponding vector $(u_1(h), \ldots, u_n(h))$. This vector consists of the payoffs for every player in case that the combination of actions required to reach that terminal node is eventually being played.

The action chosen by a player for every history after which it is his turn to move, is determined by her *strategy* function. The strategy function is defined for *all histories, even ones that would not be reached if the strategy is followed.*

Definition A.2 (Strategy)**.** *A (pure)* strategy *for player $i \in N$ in the game (N, H, P, u) is a function that assigns an action in $A(h)$ to each nonterminal history $h \in H \setminus Z$ for which $P(h) = i$.*

Note that we only consider pure (i.e. deterministic) strategies. While not all games have pure Nash equilibria, the game we construct will in fact have such an equilibrium, and thus we do not need to consider mixed (i.e. randomized) strategies. Indeed, for all the solution concepts we consider, an equilibrium with respect to pure-strategy deviations is also an equilibrium with respect to mixed-strategy deviations.

We let s_i denote the strategy employed by player i, and let $s = (s_1, \ldots, s_n)$ denote the vector of players' strategies (a.k.a. the *strategy profile*). Given a strategy vector $s = (s_1, \ldots, s_n)$ we define the outcome $o(s)$ of s to be the terminal history $h \in Z$ that results when each player $i \in N$ follows the actions chosen by s_i. In particular, $o(s)$ is a history $h = (a_1, \ldots, a_\ell)$ such that for $j = 1, \ldots, \ell - 1$ we have $s_{P(a_1,\ldots,a_j)}(a_1, \ldots, a_j) = a_{j+1}$.

The value of player i's utility under strategy vector s is equal to $u_i(o(s))$. For simplicity, it is denoted by $u_i(s)$. We assume that rational players wish to maximize this value. The equilibrium properties of our protocol will, however, only require ordinal assumptions about the preferences of players regarding different outcomes.

A.2 Subgame Perfect Equilibrium

The most basic goal is to design protocols whose prescribed strategy vector s corresponds to a *Nash equilibrium* (for the induced extensive form game). Let $(s'_i, s_{-i}) = (s_1, \ldots, s_{i-1}, s'_i, s_{i+1}, \ldots, s_n)$; that is, (s'_i, s_{-i}) denotes the strategy vector s with i's strategy changed to s'_i.

Definition A.3 (Nash equilibrium)**.** *A vector of strategies s in the game $\Gamma = (N, H, P, u)$ is said to be a* Nash equilibrium *if given that all other players are following s_{-i}, there is no incentive for player i to deviate and follow any strategy other than s_i; That is, for all $i \in N$, and for any $s'_i \neq s_i$, it holds that $u_i(s'_i, s_{-i}) \leq u_i(s)$.*

The basic notion of Nash equilibrium turns out to be an unsatisfactory solution concept for extensive-form games. The reason is that a Nash equilibrium can rely on incredible threats by players, which exist to maintain the equilibrium but never occur during the equilibrium play (and would not be in the self-interest of the player if tested.)

A more appealing solution concept is that of *subgame perfect equilibrium*. This is a strengthening of the notion of Nash equilibrium in that it requires that the equilibrium strategy is a Nash equilibrium in every *subgame* of the original extensive game.

Definition A.4 (Subgame)**.** *The* subgame *of $\Gamma = (N, H, P, u)$ that follows a history $h \in H$ is the extensive game $\Gamma(h) = (N, H|_h, P|_h, u|_h)$, where $H|_h$ is the set of sequences h' of actions for which*

$(h, h') \in H$ *(that is,* $H|_h = \{h' : (h, h') \in H\}$*), the function* $P|_h$ *is defined by* $P|_h(h') = P(h, h')$ *for each* $h' \in H_h$*, and for every* $i \in N$*, the function* $u_i|_h$ *is defined by* $u_i|_h(h') = u_i(h, h')$.[5]

Given a strategy s_i of player i and a history h in the extensive game Γ, denote by $s_i|_h$ the strategy that s_i induces in the subgame $\Gamma(h)$. That is, $s_i|_h(h') = s_i(h, h')$ for each $h' \in H|_h$. We denote by $o|_h$ the outcome function of $\Gamma(h)$, and let $u_i(s|_h)$ be the value of player i's utility under strategy vector $s|_h$ in the game $\Gamma(h)$.

Definition A.5 (Subgame perfect equilibrium)**.** *A vector of strategies* s *in the game* $\Gamma = (N, H, P, u)$ *is said to be a* subgame perfect equilibrium *if for every* $h \in H$ *the vector* $s|_h$ *is a Nash equilibrium for the game* $\Gamma(h)$*. That is, for all* $i \in N$*, every non-terminal history* $h \in H \setminus Z$ *for which* $P(h) = i$*, and any* $s'_i|_h \neq s_i|_h$*, it holds that* $u_i(s'_i|_h, s_{-i}|_h) \leq u_i(s|_h)$*.*

In principle, in order to verify that a strategy vector s is a subgame perfect equilibrium we have to check that for every player i and every subgame, there is no strategy that leads to an outcome that player i prefers. The following lemma states that we can restrict our attention, for each player i and each subgame, to alternative strategies that differ from s_i only in the actions they prescribe after *just* the initial history of this subgame.

Lemma A.6 (The one deviation property)**.** *Let* $\Gamma = (N, H, P, u)$ *be a (finite) extensive game with perfect information. The strategy vector* s *is a subgame perfect equilibrium of* Γ *if and only if for all* $i \in N$*, every* $h \in H \setminus Z$ *for which* $P(h) = i$*, and for any* $s'_i|_h$ *that differs from* $s_i|_h$ *only in the action it prescribes after the initial history of* $\Gamma(h)$*, it holds that* $u_i(s'_i|_h, s_{-i}|_h) \leq u_i(s|_h)$*.*

In other words, a strategy vector is a subgame perfect equilibrium if and only if for each subgame the player who makes the first move cannot obtain a better outcome by only changing his initial action in the corresponding subgame. We later establish that our protocol is also robust against coalitional deviations but delay the definition of the appropriate game-theoretic concepts for equilibrium given the possibility of coordination until then.

B Missing Proofs and Definitions

B.1 Impossibility if Everybody is Rational

Proposition 3.1 *Let* s *be a subgame perfect equilibrium for the game* Γ^π*, Then,* $\mathsf{num}(o) = 0$*.*

Proof: Let $h = (a_1, \ldots, a_{\ell-1}) \in H$, and let $\mathsf{rev}(h)$ denote the number of `REVEAL` actions in h. That is, $\mathsf{rev}(h)$ is the number of $j \leq \ell - 1$ for which $a_j =$ `REVEAL`.

Lemma B.1. *Let* s *be a subgame perfect equilibrium for the game* Γ^π*, let* $h \in H$ *so that* $\mathsf{rev}(h) = t-1$ *and let* $i = P^\pi(h)$*. Then,* $s_i(h) =$ `ABORT`*.*

Proof: The proof proceeds by backward induction on the length of h. Let $h = (a_1, \ldots, a_{\ell-1})$. The base case is to consider histories h of length $\ell - 1 = n - 1$ (this is where we use the fact that the number of rounds in the game Γ^π is bounded). In this case, we have that player $i = P^\pi(h)$ already knows t shares ($t - 1$ in h and the one player i is yet to reveal). In particular, no matter what is the action taken by player i, the outcome o of any possible subtree rooted at h satisfies $\delta_i(o) = 1$. Notice that if player i sends an `ABORT` message then there will exist a $j \in [\ell - 1]$ for

[5] When viewing the extensive game as a tree, a subgame that follows a history h corresponds to the subtree of the game that is rooted at node h.

which $\delta_{\pi(j)}(o) = 0$ (this will be one of the $t-1$ players that has broadcasted its values in h). On the other hand, if player i sends a `REVEAL` message then for all $j \in N$ it will hold that $\delta_j(o) = 1$ (since everybody would learn t shares). By our assumptions on the utility functions it will then hold that player i *strictly* prefers sending an `ABORT` message over a `REVEAL` message. Thus, if s is a subgame perfect equilibrium then it must be the case that $s_i(h) = \texttt{ABORT}$.

Consider now a history h of length $\ell - 1 < n$. By our induction hypothesis, for any history $h' = (h, h'')$ (of longer length) it holds that $s_{P(h')}(h') = \texttt{ABORT}$. Using the same argument as in the base case, we are then in a situation where if player i sends an `ABORT` message then there will exist a $j \in [\ell - 1]$ for which $\delta_{\pi(j)}(o) = 0$ (as in equilibrium all future players will send an `ABORT` message). In addition, if player i sends a `REVEAL` message then for all $j \in N$ it will hold that $\delta_j(o) = 1$. By our assumptions on the utility functions it will then hold that player i *strictly* prefers sending an `ABORT` message over a `REVEAL` message. As before, if s is a subgame perfect equilibrium then it must be the case that $s_i(h) = \texttt{ABORT}$. ■

By the above lemma, we may infer that any SPE strategy s must abort if exactly $t-1$ reveals took place in the history. Since the number of reveals increases by at most 1 per turn/stage, it is impossible achieve t or more reveals at the end. Thus, for any outcome $o = o(s)$ and any $i \in N$ it will hold that $\delta_i(o) = 1$, and thus $\mathsf{num}(o) = 0$. ■

B.2 Restricted Games

Definition B.2 (Restricted game)**.** *Let $\Gamma = (N, H, P, \vec{u})$ be an extensive form game, let $K \subset N$, and let s be a strategy vector. The* restricted game*, $\Gamma_{s_{-K}}$, is the extensive game that results from playing Γ with players $i \in N \setminus K$ always following the strategy s_i; i.e., $\Gamma_{s_{-K}} = (K, H_{s_{-K}}, P_{s_{-K}}, u_K)$, where $H_{s_{-K}}$ is the set of sequences $h = (a_1, \ldots, a_\ell) \in H$ so that for all $P(a_1, \ldots, a_j) \in N \setminus K$, it holds that $a_{j+1} = s_{P(a_1,\ldots,a_j)}(a_1, \ldots, a_j)$, and $P_{s_{-K}}$ is the function defined by*

$$P_{s_{-K}}(h) = \begin{cases} P(h) & \text{if } P(h) \in K \\ P_{s_{-K}}(h, s_{P(h)}(h)) & \text{if } P(h) \in N \setminus K \end{cases}$$

We view the players in the set K as being rational, and thus are still free to deviate from the strategy vector s, where the players in $N \setminus K$ are honest and will always follow s. We will show there are enough honest players (but still a significant minority), then protocol $(\Gamma^{\pi}_{\mathbf{s}_{-K}}, \mathbf{s}_K)$ will be in (subgame perfect) equilibrium for the vast majority of permutations π. Notice that in the protocol $(\Gamma^{\pi}_{\mathbf{s}_{-K}}, \mathbf{s}_K)$, the prescribed instructions are identical for all players, honest or otherwise. This is useful because one can motivate the existence of a small number of honest players by assuming that some participants in a protocol are too disinterested, uninformed, or simply lazy to deviate from the prescribed strategy (in addition to those who may be honest out of altruism).

B.3 Subgame Perfect Equilibrium

Theorem 4.1 (Subgame perfect equilibrium) *Let $K \subset N$ and suppose π is a K-good permutation. Then, the strategy vector $\mathbf{s}_K$ is a subgame perfect equilibrium for the game $\Gamma^{\pi}_{\mathbf{s}_{-K}}$.*

Proof: Fix a set $K \subset N$ and a K-good permutation π. By assumption, the players $k \in N \setminus K$ always follow the prescribed strategy. Our goal is to show that the strategy vector $\mathbf{s}_K$ is a subgame perfect equilibrium for the restricted game $\bar{\Gamma} = \Gamma^{\pi}_{\mathbf{s}_{-K}}$ (for simplicity, we denote $\mathbf{s}_K$ by $\mathbf{s}$). Let $\bar{\Gamma} = (K, H, P, u)$. By Lemma A.6, for any player $i \in K$ and every subgame $\bar{\Gamma}(h)$ rooted at h so

that $P(h) = i$, it will be sufficient to focus on comparing the payoffs i receives by playing $\mathbf{s}$ versus playing (s'_i, s_{-i}) where $s'_i(h) \neq s_i(h)$.[6]

Let $h = (a_1, \ldots, a_{\ell-1})$ be the root of the subtree, and let $\mathsf{rev}(h)$ denote the number of `REVEAL` actions in h. That is, $\mathsf{rev}(h)$ is the number of $j \leq \ell - 1$ for which $a_j = \texttt{REVEAL}$. We start by analyzing the case in which $\mathsf{rev}(h) \geq t$.

Lemma B.3. *Suppose that* $\mathsf{rev}(h) \geq t$. *Then,* $u_i(\texttt{REVEAL}, \mathbf{s}_{-i}) = u_i(\texttt{ABORT}, \mathbf{s}_{-i})$.

Proof: By the time that it is the turn of player i to take action, the secret is already known to all players. Thus, *no matter* what is the action taken by i, the outcome o of any possible eventual subgame in the subtree rooted at h will satisfy $\delta_j(o) = 1$ for *all* $j \in [n]$. From our assumptions on the utility functions, we would then have that player i is *indifferent* between outcomes that result from him playing `REVEAL` and outcomes that result from him playing `ABORT`; i.e., $u_i(\texttt{REVEAL}, \mathbf{s}_{-i}) = u_i(\texttt{ABORT}, \mathbf{s}_{-i})$. ∎

¿From this point on, we may thus assume that $\mathsf{rev}(h) < t$. Our analysis distinguishes between two cases, depending on whether one of the first $t - 1$ (or $\ell - 1$, in case that $\ell < t$) actions in the history is `ABORT`. More specifically, let

$$t^* \stackrel{\text{def}}{=} \min\{\ell, t\}.$$

We distinguish between the case that there exist some $j < t^*$ for which $a_j = \texttt{ABORT}$, and the case that $a_j = \texttt{REVEAL}$ for all $j \leq t^*$. Each of these cases will require a separate treatment, since in the former case the prescribed strategy instructs player i to send an `ABORT` message, whereas in the latter it instructs him to send a `REVEAL` message.

The following lemma deals with the first case, in which $a_j = \texttt{ABORT}$ for some $j < t^*$. By the definition of the prescribed strategy s_i, we have $s_i(h) = \texttt{ABORT}$. Thus, what we will need to argue is that, assuming all players $\pi(k)$ for $k > \ell$ follow their prescribed strategy, player $i = \pi(\ell)$ prefers following $s_i(h) = \texttt{ABORT}$ over the (only possible) alternative strategy $s'_i(h) = \texttt{REVEAL}$.

Lemma B.4. *Suppose that* $a_j = \texttt{ABORT}$ *for some* $j < t^*$. *Then,* $u_i(\texttt{REVEAL}, \mathbf{s}_{-i}) \leq u_i(\texttt{ABORT}, \mathbf{s}_{-i})$.

Proof: We distinguish between the following two cases, depending on the value of $\mathsf{rev}(h)$:

- $\mathsf{rev}(h) < t - 1$: Since there exists $j < t^*$ for which $a_j = \texttt{ABORT}$ and since $t^* \leq t$, then for all $k \geq \ell$, the prescribed action for player $\pi(k)$ is to send an `ABORT` message. Thus, no matter what is the action taken by player i, if all subsequent players $\pi(k)$ follow their prescribed strategies, the total number of shares that are eventually known by player i will be at most $\mathsf{rev}(h) < t - 1$. This means that for either actions available to player i, the outcome o of the game will satisfy $\delta_i(o) = 0$. By our assumptions on the utility functions, it then follows that player i is indifferent between outcomes that result from him playing `REVEAL` and outcomes that result from him playing `ABORT`; i.e., $u_i(\texttt{REVEAL}, \mathbf{s}_{-i}) \leq u_i(\texttt{ABORT}, \mathbf{s}_{-i})$.[7]

- $\mathsf{rev}(h) = t - 1$: Since at the node in which the subtree is rooted player i is yet to broadcast its value, we have that player i knows t shares. Thus, no matter what action player i chooses, every possible outcome o in the subtree will satisfy $\delta_i(o) = 1$. Now, since $a_j = \texttt{ABORT}$ for

[6] Since every player in $\bar{\Gamma}$ plays exactly once, and since in every level of the game tree there is only one player that is supposed to play, we may omit the restriction $|_h$ from $s_i|_h$. This results in the simpler notation s_i.

[7] Note that if player i sends a `REVEAL` message then it is possible that for some $k \neq i$ it will hold that $\delta_k(o) = 1$ (while $\delta_i(o) = 0$ no matter how player i acts). Thus, in some cases, player i might *strictly* prefer `ABORT` over `REVEAL`.

some $j < t^*$, we have that for all $k \geq \ell$, the prescribed action for player $\pi(k)$ is to send an ABORT message. Thus, if the action taken by player i is ABORT, and all subsequent players $\pi(k)$ follow their prescribed strategies, there will exist players $m \in [n]$ that will eventually learn at most $t-1$ shares and for which $\delta_m(o) = 0$. (These will be the players $m = \pi^{-1}(j)$, who correspond to the $t-1$ indices $j \leq \ell$ for which $a_j = \texttt{REVEAL}$.) If, on the other hand, the action taken by player i is REVEAL, then we would have that $\delta_m(o) = 1$ for all $m \in [n]$. By our assumptions on the utility functions, it then follows that player i *strictly* prefers outcomes that result from him playing ABORT over outcomes that result from him playing REVEAL; i.e., $u_i(\texttt{REVEAL}, \mathbf{s}_{-i}) \leq u_i(\texttt{ABORT}, \mathbf{s}_{-i})$.

This completes the proof of Lemma B.4. ■

We now turn to handle the case in which $a_j = \texttt{REVEAL}$ for all $j \leq t^*$. By the definition of the prescribed strategy s_i, we have $s_i(h) = \texttt{REVEAL}$. Thus, what we will need to argue is that, assuming all players $\pi(k)$ for $k > \ell$ follow their prescribed strategy, player $i = \pi(\ell)$ prefers sending a REVEAL message over sending an ABORT message.

Lemma B.5. *Suppose that π is K-good, and that $a_j = \texttt{REVEAL}$ for all indices $j < t^*$. Then, $u_i(\texttt{ABORT}, \mathbf{s}_{-i}) \leq u_i(\texttt{REVEAL}, \mathbf{s}_{-i})$.*

Proof: We distinguish between the following two cases, depending on the value of $\mathsf{rev}(h)$:

$\mathsf{rev}(h) < t-1$: Since $a_j = \texttt{REVEAL}$ for all indices $j < t^*$, then the prescribed action for player i is to send a REVEAL message. Thus, if player i indeed chooses to follow the prescribed strategy an send REVEAL, then for all $k > \ell$ the prescribed action for a subsequent player $\pi(k)$ will be to send a REVEAL message. This would entail that the number of shares eventually revealed is n and that $\delta_k(o) = 1$ for all $k \in [n]$ (and in particular $\delta_i(o) = 1$).

On the other hand, if player i chooses to send an ABORT message, then the prescribed action for subsequent players $\pi(k)$ will be to send an ABORT message, resulting in $\delta_i(o) = 0$. By our assumptions on the utility functions, player i *strictly* prefers outcomes o for which $\delta_i(o) = 1$ over outcomes o' for which $\delta_i(o') = 0$. Thus, player i will strictly prefer to take the action REVEAL over the action ABORT; i.e., $u_i(\texttt{ABORT}, \mathbf{s}_{-i}) \leq u_i(\texttt{REVEAL}, \mathbf{s}_{-i})$.

$\mathsf{rev}(h) = t-1$: This is the place where we will be using our hypothesis that the permutation π is K-good and that players $k \in N \setminus K$ always act as prescribed. Since $\mathsf{rev}(h) = t-1$, and since at the node in which the subtree is rooted player i is yet to broadcast its value, we have that player i knows t shares. Thus, no matter what action player i chooses, every possible outcome o in the subtree will satisfy $\delta_i(o) = 1$. We will next need to argue that player i is indifferent between REVEAL and ABORT. Given that $\delta_i(o) = 1$, this can happen only if player i's action *has no effect* on the value of $\delta_k(o)$ for any $k \neq i$. Here we use the fact that the permutation π is K-good.

Recall that K-goodness implies the existence of $k \geq t$ so that player $\pi(k) \in N \setminus K$, which by definition of the restricted game $\bar{\Gamma} = \Gamma^{\pi}_{\mathbf{s}_{-K}}$, implies that player $\pi(k)$ always follows the prescribed strategy $s_{\pi(k)}$. Since by hypothesis, $a_j = \texttt{REVEAL}$ for all $j < t^*$ and since $k \geq t \geq t^*$, then in all subtrees rooted at h the strategy $s_{\pi(k)}$ instructs player $\pi(k)$ to send a REVEAL message. As a consequence, for every terminal history o in the subtree, it will be the case that $a_k = \texttt{REVEAL}$.

Thus, no matter what is the action taken by player i, we will have that the total number of shares eventually revealed is at least t ($t-1$ in h plus at least one by player $\pi(k)$). This implies

that for any possible outcome o in the subtree it holds that $\delta_j(o) = 1$ for all $j \in [n]$. By our assumptions on the utility functions, having $\delta_j(o) = 1$ for all $j \in [n]$ regardless of how player i acts implies that he is indifferent between outcomes that result from him playing `ABORT` over outcomes that result from him playing `REVEAL`; i.e., $u_i(\texttt{ABORT}, \mathbf{s}_{-i}) \leq u_i(\texttt{REVEAL}, \mathbf{s}_{-i})$.

This completes the proof of Lemma B.4 ■

This completes the proof of Theorem 4.1. ■

B.4 PCPNE Implies SPE

Claim 5.3 *For any game Γ, if s^* is PCPNE in Γ, then s^* is SPE in Γ.*

Proof. Since strategy s^* is PCPNE in the game Γ, then it follows that s^* is PSE in Γ. Considering only single player coalitions $C = \{i\}$, we have that s_i^* is PCPNE in the restricted game $\Gamma_{s^*_{-i}}$. Hence, there does not exists another strategy s_i' such that $u_i(s_i', s^*_{-i}) > u_i(s^*)$. (Otherwise, s_i' is a Pareto improvement for player i in $\Gamma_{s^*_{-i}}$.) This property also holds for all proper subgames of Γ since s^* is PCPNE in all proper subgames of Γ, which follows from s^* being PSE in Γ. Therefore, s^* is SPE in the game Γ. ■

B.5 Proof of PCPNE

Theorem 5.4 *Let protocol $(\Gamma^{\pi}_{\mathbf{s}_{-K}}, \mathbf{s}_K)$ be defined as in Definition 2.2, with π being a K-good permutation and $\mathbf{s}$ being our prescribed strategy. Then, strategy $\mathbf{s}_K$ is PCPNE in the game $\Gamma^{\pi}_{\mathbf{s}_{-K}}$.*

Let the game $\Gamma^{\pi}_{\mathbf{s}_{-K}}$ be denoted by $\overline{\Gamma} = (K, \overline{H}, \overline{P}, \overline{u})$. To prove the above theorem, we characterize all PCPNE strategies for the game $\overline{\Gamma}$ by defining *good strategies* as follows: For a history $h \in \overline{H}$ and a set of players $C \subseteq K$, we say strategy s is *good with respect to C and h* iff for every history $h' \in \overline{H}|_h$, letting $i = P(h')$, the following conditions are satisfied.

1. If $i \in C$, $\mathsf{err}(h') = \texttt{yes}$, and $t - 2 \leq \mathsf{rev}(h') \leq t - 1$, then $s_i(h') = \texttt{abort}$.
2. If $i \in C$, $\mathsf{err}(h') = \texttt{no}$ and $\mathsf{rev}(h') < t - 1$, then $s_i(h') = \texttt{reveal}$. (Note that this implies that we are at the $(\mathsf{rev}(h') + 1)$-th stage of the protocol.)
3. Otherwise, in all other cases, $s_i(h')$ is unconstrained, i.e., it can be either `reveal` or `abort`.

Recall that $\mathsf{err}(h')$ denotes whether an abort occurred in the first $t-1$ rounds of h', and $\mathsf{rev}(h'')$ is the number of reveals that took place in h'.

We establish four claims concerning the properties of good strategies. The first claim follows from definition of the prescribed strategy $\mathbf{s}$; see Section 2.2.

Claim B.6. *The prescribed strategy $\mathbf{s}_K$ is good with respect to all the rational players K and the null history $h = \varepsilon$.*

The second claim follows from the observation that whether a strategy s is good with respect to C and h involves checking every node of the game tree after h in which a player in C moves; hence a good strategy is "local" property.

Claim B.7. *For any two sets C' and C such that $C' \subseteq C \subseteq K$, and any history $h \in \overline{H}$, the following two conditions are equivalent.*

1. *Strategy s is good with respect to C and h.*

2. *Strategy s is good with respect to C' and h, and s is good with respect to $C \setminus C'$ and h.*

The third claim states that good strategies are "optimal," in the sense that not playing good strategies gives strictly inferior outcomes assuming everyone else plays good strategies.

Claim B.8. *For any history $h \in \overline{H}$, letting $i = P(h)$ and $C = K \setminus \{i\}$, if s that is good with respect to K and h but s' is* not *good with respect to $\{i\}$ and h, then $\overline{u}_i((s_i, s_{-i})|_h) > \overline{u}_i((s'_i, s_{-i})|_h)$.*

The final claim states that any two good strategies are "equally as good" for all the rational players in our protocol.

Claim B.9. *For every history $h \in \overline{H}$, and any two good strategies s and s' with respect to K and h, we have $u_i(s|_h) = u_i(s|_h)$ for all $i \in K$.*

Proof. We consider every node in the game tree, with each node represented by a history h, and argue that playing any two good strategies $s|_h$ and $s'|_h$ would end up with $u_i(s|_h) = u_i(s|_h)$ for all $i \in K$.

Consider the case when $\mathsf{err}(h) = \texttt{yes}$. Then, $s_i(h)$ is unconstrained only if $\mathsf{rev}(h) \geq t$ or $\mathsf{rev}(h) < t - 2$. In the former case, when $\mathsf{rev}(h) \geq t$, it does not matter whether player i reveals or aborts since every player would have learnt the secret. In the latter case, when $\mathsf{rev}(h) < t - 2$, $s_i(h) = \texttt{reveal}$ would increase the number of revealed shares by one, but as soon as the number of revealed shares reaches $t - 2$, the succeeding players would all abort (by condition 1 above), resulting in an outcome where no player learns the secret. On the other hand, if $s_i(h) = \texttt{abort}$, we would remain in the case where less than $t - 2$ shares have been revealed, and by induction, we would also be in an outcome where no player learns the secret.

Next, consider the case when $\mathsf{err}(h) = \texttt{no}$. In this case, $s_i(h)$ is unconstrained only if $\mathsf{rev}(h) \geq t - 1$. Since $\mathsf{err}(h) = \texttt{no}$ and permutation π is K-good, there would be an honest player after the t-th stage who would reveal her share. Therefore, everyone will learn the secret whether or not player i reveals her share. ■

The next lemma shows that good strategies exactly characterizes the set of PCPNE strategies in our game.

Lemma B.10. *For every history $h \in \overline{H}$ and every set $C \subseteq K$, the following three conditions are equivalent for strategies s that are good with respect to $K \setminus C$ and h.*

1. *s is good with respect to C and h;*

2. *$s_C|_h$ is PSE in $\overline{\Gamma}_{s_{-C}}|_h$;*

3. *$s_C|_h$ is PCPNE in $\overline{\Gamma}_{s_{-C}}|_h$.*

Theorem 5.4 follows from Lemma B.10 by taking $C = K$ and $h = \varepsilon$, and by noting that the prescribed strategy $\mathbf{s}$ is good with respect to K and $h = \varepsilon$, by Claim B.6.

Proof. We prove the equivalence by induction on the number of players n and number of stages r in the game $\overline{\Gamma}_{s_{-C}}|_h$. Note that $n \leq |C|$ and $r \leq |K| - |h|$.

The base case where $(n, r) = (1, 1)$ is handled Claim B.8, which states that playing good strategies is "optimal" assuming every other subsequent rational player plays good strategies.

We now prove the inductive step. From now on, let game $\Gamma' = (C, H', P', u')$ denote the game $\overline{\Gamma}_{s_{-C}}|_h$.

(1) $\Rightarrow$ (2) Since s is good with respect to C and h, and s that are good with respect to $K \setminus C$ and h, then by Claim B.7, s is good with respect to K and h.

For any $h' \in H' \setminus \{\varepsilon\}$ and $C' \subsetneq C$, applying Claim B.7 again gives us: (i) s' is good with respect to C and h', and s' is good with respect to $K \setminus C$ and h'; (ii) s' is good with respect to C' and ε, and s' is good with respect to $K \setminus C'$ and ε. Then by induction, s' is PCPNE in all proper subgames of Γ', and also PCPNE in all proper coalitions $C' \subsetneq C$ of Γ'. This implies that s' is PSE in Γ'.

(2) $\Rightarrow$ (1) Assume for sake of contradiction that s' is PSE in Γ' but *not* good with respect to C and h. By the definition of PSE, s' is PCPNE in every proper subgame of Γ', and by induction s' is good with respect to C and any $h' \in H'$.

Consequently, it is the move of the immediate next player $i = P'(h')$, namely $s'_i(h')$, that contradicts the requirement of being good. In other words, s' is *not* good with respect to $\{i\}$ and h'. By Claim B.8, the utility of player i when playing s'_i is strictly inferior to playing a good strategy s''_i. This contradicts the fact that s' is PSE in Γ'.

(2) $\Rightarrow$ (3) Having established the equivalence between good and PSE strategies, namely (1) $\Leftrightarrow$ (2), we use Claim B.9 to conclude that the choice of good (or equivalently, PSE) strategies in the game Γ' does not affect the utilities of the rational players. This implies that no PSE strategy in Γ' can Pareto improve any other PSE strategy. Hence, every PSE strategy is also PCPNE.

(3) $\Rightarrow$ (2) This direction is immediate since PCPNE strategies must be PSE; see Definition 5.2.

■

B.6 Comments

Remark B.1. *It may be helpful to contrast this approach with that of Lepinski et al. [23] who are also concerned about coalitional deviations in the design of distributed protocols. Specifically, they adopt cryptographic methods to ensure that there are no opportunities for coordinating on deviations by sending messages* ***within*** *the message space of a protocol. The view adopted in our work is quite different. The incentives in our protocol are naturally aligned such that even if (as in the standard game theory story) the players are able to enter "smoke filled rooms" they can nevertheless not agree on a useful deviation that will be stable against further deviations.*

Remark B.2. *In considering coalitional deviations we are entertaining the possibility that parties in a protocol can communicate freely outside of the protocol in forming agreements about play within the protocol. This is the standard game theory story for coalitional-proof Nash equilibrium, i.e. even if groups of players can discuss in "cheap talk" how to play they will be unable to identify useful deviations. Why, one might ask, would a coalition of players not simply proceed to share their secret outside of the protocol. We give a response in two parts. (1) For concreteness consider the following scenario. The protocol is executed in a room in which every party has his share of the secret inside a sealed envelope on a table. The available actions of a party are to* reveal *(and open the envelope) or* abort *(refuses to open the envelope). To consider coalitional deviations we also allow players to freely leave the room in any round, but without the envelope, and engage in discussion. The envelopes– and the shares –remain in the room. (2) Now, any subset of the players may nevertheless choose to abort and leave the room with their envelopes (and unable now to return to the room). But, if we assume neither a mediator nor a simultaneous broadcast (or similar)*

device outside the room that it seems incredible, for the same reasons as the impossibility results for rational secret sharing [22, 20] for these players to be able to coordinate amongst themselves such that they all learn the secret! Thus, we see that the players want to use the designed protocol (in the room) with the small minority of honest players because it is precisely this protocol that allows for secret sharing. One can also make the same argument for a different concrete setting in which players can freely leave with their envelope and return. Again, while players may choose to engage in "cheap talk" outside of the room about how to play inside the room it seems reasonable to assume they will not begin to open envelopes outside of the room.